



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2017

---

## **Randomized trials addressing a similar question are commonly published after a trial stopped early for benefit**

Murad, M Hassan ; Guyatt, Gordon H ; Domecq, Juan Pablo ; Vernooij, Robin W M ; Erwin, Patricia J ; Meerpohl, Joerg J ; Prutsky, Gabriela J ; Akl, Elie A ; Mueller, Katharina ; Bassler, Dirk ; Schandelmaier, Stefan ; Walter, Stephen D ; Busse, Jason W ; Kasenda, Benjamin ; Pagano, Gennaro ; Pardo-Hernandez, Hector ; Montori, Victor M ; Wang, Zhen ; Briel, Matthias

**Abstract:** **OBJECTIVE** We explored how investigators of ongoing or planned trials respond to the publication of a trial stopped early for benefit addressing a similar question. **STUDY DESIGN AND SETTING** We searched multiple databases from the date of publication of the truncated trial through August, 2015. Independent reviewers selected trials and extracted data. **RESULTS** We identified 207 trials truncated for early benefit; of which 102 (49%) were followed by subsequent trials (262 subsequent trials, median 2 per truncated trial, range 1-13). Only 99 (38%) provided a rationale justifying conducting a trial despite prior stopping. The top reasons were to address different population or setting (33%); skepticism of truncated trials findings because of small sample size (12%), inconsistency with other evidence (11%) or increased risk of bias (7%). We did not identify significant associations between subsequent trials and characteristics of truncated ones (risk of bias, precision, funding, or rigor of stopping decision). **CONCLUSION** About half of the trials stopped early for benefit were followed by subsequent trials addressing a similar question. This suggests that future trialists may have been skeptic about the decision to stop prior trials. A more rigorous threshold for stopping early for benefit is needed.

DOI: <https://doi.org/10.1016/j.jclinepi.2016.10.006>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-129806>

Journal Article

Accepted Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

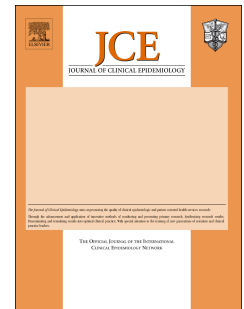
Murad, M Hassan; Guyatt, Gordon H; Domecq, Juan Pablo; Vernooij, Robin W M; Erwin, Patricia J; Meerpohl, Joerg J; Prutsky, Gabriela J; Akl, Elie A; Mueller, Katharina; Bassler, Dirk; Schandelmaier, Stefan; Walter, Stephen D; Busse, Jason W; Kasenda, Benjamin; Pagano, Gennaro; Pardo-Hernandez, Hector; Montori, Victor M; Wang, Zhen; Briel, Matthias (2017). Randomized trials addressing a similar question are commonly published after a trial stopped early for benefit. *Journal of Clinical Epidemiology*, 82:12-19.

DOI: <https://doi.org/10.1016/j.jclinepi.2016.10.006>

# Accepted Manuscript

Randomized trials addressing a similar question are commonly published after a trial stopped early for benefit

M. Hassan Murad, Gordon H. Guyatt, MD, MSc, Juan Pablo Domecq, MD, Robin W.M. Vernooij, Patricia J. Erwin, Joerg J. Meerpohl, MD, Gabriela J. Prutsky, MD, Elie A. Akl, Katharina Mueller, MD, Dirk Bassler, Stefan Schandelmaier, MD, Stephen D. Walter, PhD, Jason W. Busse, DC, PhD, Benjamin Kasenda, MD, PhD, Gennaro Pagano, MD, MSc, Hector Pardo-Hernandez, BA, MPH, PhD, Victor M. Montori, Zhen Wang, PhD, Matthias Briel



PII: S0895-4356(16)30585-6

DOI: [10.1016/j.jclinepi.2016.10.006](https://doi.org/10.1016/j.jclinepi.2016.10.006)

Reference: JCE 9263

To appear in: *Journal of Clinical Epidemiology*

Received Date: 2 February 2016

Revised Date: 10 September 2016

Accepted Date: 1 October 2016

Please cite this article as: Murad MH, Guyatt GH, Domecq JP, Vernooij RWM, Erwin PJ, Meerpohl JJ, Prutsky GJ, Akl EA, Mueller K, Bassler D, Schandelmaier S, Walter SD, Busse JW, Kasenda B, Pagano G, Pardo-Hernandez H, Montori VM, Wang Z, Briel M, Randomized trials addressing a similar question are commonly published after a trial stopped early for benefit, *Journal of Clinical Epidemiology* (2016), doi: 10.1016/j.jclinepi.2016.10.006.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Randomized trials addressing a similar question are commonly published after a trial stopped early for benefit

### Authors and affiliation

M. Hassan Murad [murad.mohammad@mayo.edu](mailto:murad.mohammad@mayo.edu)  
 Evidence-based Practice Center, Mayo Clinic, Rochester, MN, USA  
 Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN, USA  
 Division of Preventive Medicine, Mayo Clinic, Rochester, MN, USA

Gordon H. Guyatt, MD, MSc [guyatt@mcmaster.ca](mailto:guyatt@mcmaster.ca)  
 Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON L8S 4L8, Canada.

Juan Pablo Domecq, MD [JDOMEQCQ1@hfhs.org](mailto:JDOMEQCQ1@hfhs.org)  
 1. Henry Ford Hospital, Department of Internal Medicine, Detroit, MI  
 2. Unidad de Conocimiento y Evidencia, CONEVID, UPCH, Lima, Peru  
 3. KER Unit

Robin W.M. Vernooij [robinvernooij@gmail.com](mailto:robinvernooij@gmail.com)  
 Iberoamerican Cochrane Centre, Institute of Biomedical Research (IIB Sant Pau), Barcelona, Spain.

Patricia J Erwin [Erwin.patricia@mayo.edu](mailto:Erwin.patricia@mayo.edu)  
 Mayo Clinic Libraries, Mayo Clinic, Rochester, MN, USA

Joerg J. Meerpohl, MD [meerpohl@cochrane.de](mailto:meerpohl@cochrane.de)  
 Cochrane Germany, Medical Center – University of Freiburg, Berliner Allee 29, 79110 Freiburg, Germany

Gabriela J Prutsky, MD [gapru24@gmail.com](mailto:gapru24@gmail.com)  
 1. Children's Hospital of Michigan, Department of Pediatrics, Detroit, Michigan  
 2. Unidad de Conocimiento y Evidencia, CONEVID, UPCH, Lima, Peru  
 3. KER Unit

Elie A. Akl [ea32@aub.edu.lb](mailto:ea32@aub.edu.lb)  
 Department of Internal Medicine, American University of Beirut, Beirut, Lebanon  
 Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada

Katharina Mueller, MD [Katharina-F.Mueller@gmx.de](mailto:Katharina-F.Mueller@gmx.de)

Center for Clinical Pediatric Studies, University Children's Hospital Tuebingen,  
Frondsbergstraße 23, 72070 Tuebingen, Germany

Dirk Bassler [dirk.bassler@usz.ch](mailto:dirk.bassler@usz.ch)  
Professor (or MD, MSc), Head of Department, Department of Neonatology, University  
Hospital Zurich and University of Zurich, Frauenklinikstrasse 10, 8091 Zurich,  
Switzerland

Stefan Schandelmaier, MD [stefan.schandelmaier@usb.ch](mailto:stefan.schandelmaier@usb.ch)  
Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel,  
Switzerland

Stephen D. Walter, PhD [walter@mcmaster.ca](mailto:walter@mcmaster.ca)  
Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton,  
ON L8S 4L8, Canada.

Jason W. Busse, DC, PhD [bussejw@mcmaster.ca](mailto:bussejw@mcmaster.ca)  
The Michael G. DeGroote Institute for Pain Research and Care, Department of  
Anesthesia, and Department of Clinical Epidemiology and Biostatistics, McMaster  
University, Hamilton, Ontario, Canada

Benjamin Kasenda, MD, PhD [benjamin.kasenda@gmail.com](mailto:benjamin.kasenda@gmail.com)  
1. Basel Institute for Clinical Epidemiology  
2. Royal Marsden Hospital, Dep Medicine, London, UK

Gennaro Pagano [gennaro.pagano@unina.it](mailto:gennaro.pagano@unina.it)  
Gennaro Pagano, MD, MSc. Research Scientist. Federico II University of Naples.  
Naples, Italy.

Hector Pardo-Hernandez, BA, MPH, PhD [hpardo@santpau.cat](mailto:hpardo@santpau.cat)  
Iberoamerican Cochrane Centre, Institute of Biomedical Research (IIB Sant Pau),  
Barcelona, Spain.

Victor M Montori [montori.victor@mayo.edu](mailto:montori.victor@mayo.edu)  
Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN, USA

Zhen Wang, PhD [Wang.Zhen@mayo.edu](mailto:Wang.Zhen@mayo.edu)  
Evidence-based Practice Center, Mayo Clinic, Rochester, MN, USA

Matthias Briel [matthias.briel@usb.ch](mailto:matthias.briel@usb.ch)  
1. Basel Institute for Clinical Epidemiology  
and Biostatistics, University Hospital Basel, Basel, Switzerland.  
2. Department of Clinical Epidemiology and Biostatistics, McMaster University,  
Hamilton, ON L8S 4L8, Canada.

### Acknowledgements:

Drs Murad and Briel had equal contributions to this work.

We thank these individuals for contributions in earlier stages of this project including reviewing the protocol or screening citations for inclusion: Nacho Ferreira-Gonzalez, Annette Kristiansen, Anna Silva, Lorenzo Moja, Osama Quassim Agha, M. Bassam Sonbol and Paul Glasziou.

Drs Murad will have responsibility for this publication.

The authors declare they have no relevant competing or financial interests and no financial relationships with any organizations that might have an interest in the submitted work in the previous three years and no other relationships or activities that could appear to have influenced the submitted work.

Transparency declaration: Dr Murad affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Ethical approval: Not required

Funding: None.

Contribution statement: All authors have participated sufficiently to merit authorship. The idea for the study was conceived by Briel, Guyatt, Murad, Montori and Domecq. Literature search was done by Patricia Erwin. Study selection, appraisal and data extraction was done by Murad, Briel, Domecq, Vernooj, Meerpohl, Prutsky, Akl, Busse, Mueller, Schandelmaier, Bassler, kasendra, Pagano, Pardo-Hernandez, and Wang. Statistical analysis by Wang. Comments on design and interpretation by Walter. All authors reviewed manuscript and made critical revisions and approved the final version.

## ABSTRACT

**Objective:** We explored how investigators of ongoing or planned trials respond to the publication of a trial stopped early for benefit addressing a similar question.

**Study Design and Setting:**

We searched multiple databases from the date of publication of the truncated trial through August, 2015. Independent reviewers selected trials and extracted data.

**Results:**

We identified 207 trials truncated for early benefit; of which 102 (49%) were followed by subsequent trials (262 subsequent trials, median 2 per truncated trial, range 1-13). Only 99 (38%) provided a rationale justifying conducting a trial despite prior stopping. The top reasons were to address different population or setting (33%); skepticism of truncated trials findings because of small sample size (12%), inconsistency with other evidence (11%) or increased risk of bias (7%). We did not identify significant associations between subsequent trials and characteristics of truncated ones (risk of bias, precision, funding, or rigor of stopping decision).

**Conclusion**

About half of the trials stopped early for benefit were followed by subsequent trials addressing a similar question. This suggests that future trialists may have been skeptic about the decision to stop prior trials. A more rigorous threshold for stopping early for benefit is needed.

**Key words:** Randomized controlled trials, trials stopped early for benefit, early termination of trials, systematic review, methodology, trial design

## BACKGROUND

Randomized controlled trials are sometimes stopped early by trialists if one of the interventions appears to be associated with a large benefit. Investigators may halt their trial to avoid depriving participants in the comparison group of a beneficial treatment, and to ensure rapid dissemination of the intervention.<sup>(1)</sup> Trials stopped early for benefit (truncated trials) are usually published in high-impact journals, receive considerable attention, and are likely to influence practice.<sup>(2)</sup> However; truncated trials tend to overestimate the magnitude of benefit by approximately up to one third, and overestimates can be much greater when sample sizes and number of events are modest.<sup>(3)</sup>

Further, the majority of published truncated trials fail to adequately report at least one important factor regarding the decision to stop early; such as the planned sample size, details of interim analyses, whether a stopping rule informed the decision to stop, or whether the analysis was adjusted to account for interim monitoring and truncation.<sup>(2)</sup> Misleading overestimates from truncated trials seriously threaten the integrity of decisions made by patients and clinicians when they trade off the benefits and harms of interventions. Unfortunately, the majority (71%) of systematic reviews that included truncated trials did not comment or recognize this possible bias.<sup>(4)</sup> Simulation studies have shown that when trials stopped early for benefit are included in a meta-analysis, the pooled effect size and heterogeneity parameters become distorted.<sup>(5)</sup> Therefore, this issue affects the synthesis of evidence and subsequent decision making that depends on systematic reviews, such as guidelines. For example, a trial evaluated the efficacy of bisoprolol in patients with a positive dobutamine echocardiography



undergoing elective vascular surgery. It showed that bisoprolol significantly reduced the risk of perioperative myocardial infarction and cardiac death. The trial was stopped early because of this large effect.<sup>(6)</sup> Guidelines in Europe and the U.S. recommended this intervention; which was implemented on a large scale.<sup>(7)</sup> Subsequent trials showed markedly different results and demonstrated that the reduction in myocardial infarction was not as large as originally demonstrated and that the intervention increased the risk of stroke, hypotension and may increase mortality.<sup>(7)</sup>

The motivation to stop a trial early for benefit should be balanced against the risk of disseminating overestimated treatment effect. It should also be balanced against the loss of the opportunity to generate more precise evidence and opportunity to capture the effect of treatment on secondary outcomes and outcomes that require longer follow up (particularly adverse effects). If stopping a trial early for benefit was the correct decision (ie, it would be unethical to continue the trial); then the conduct of subsequent trials addressing the same question (subsequent trials) would also be unethical.

Several justifiable reasons for launching subsequent trials are plausible. First, researchers may want to test the intervention in a population or setting that are somewhat different from that of a truncated trial. Second, researchers may be skeptical about the results of the truncated trials (because the trial was small or at high risk of bias). Third, researchers may be interested in knowing the effect of the treatment on other outcomes.

If investigators, the clinical community, and ethics committees sanction subsequent trials, it raises serious questions regarding the initial decision to truncate the

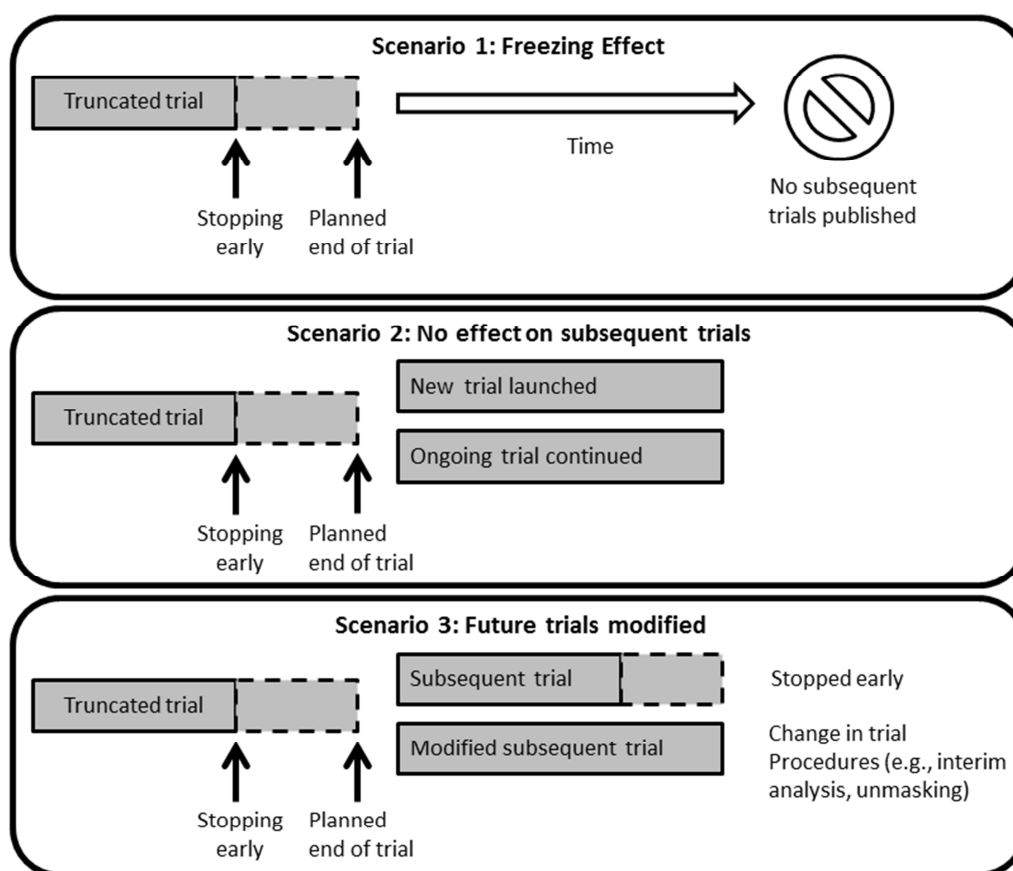
original trials for benefit. To explore the incidence of and rationale for subsequent trials, we conducted a meta-epidemiological study addressing how often subsequent trials were launched or continued after the publication of a truncated trial asking the same or similar research question.

## **METHODS/DESIGN**

The protocol of this study has been published and provides further details.<sup>(8)</sup> In brief, we identified a cohort of 207 truncated trials (published 1970-2007) <sup>(2, 3)</sup> through systematic searches of electronic databases (including MEDLINE, EMBASE, and Cochrane), communication with content experts, and manual review of journals.<sup>(3)</sup> We then identified published subsequent trials for each truncated trial. We defined a subsequent trial as a subsequent RCT that was launched (i.e., started enrollment) or continued enrollment after the truncated trial publication date and addressed a similar question (similar population, intervention, comparison and outcome). We only included subsequent trials with parallel design that continued follow-up or patient enrollment for at least 6 months after truncated trial publication.

We hypothesized 3 possible reactions of researchers (scenarios) to the publication of a truncated trial (figure). Researchers may stop conducting future similar trials (ie, truncated trial caused a freezing effect on future research) (a), launch a new trial or continue ongoing trials without reaction to truncated trial (b), or modify the protocol of an ongoing trial based on results of the truncated trial, or even stop it early (c).

**Figure: The possible effects of a trial stopped early for benefit on future research.**



### *Literature search*

With input from study investigators, a reference librarian designed and executed 207 individual search strategies, corresponding to each truncated trial, in MEDLINE, EMBASE, Cochrane Library, Web of Science, Scopus, and PsycINFO from the date of publication of the truncated trial to August, 2015. These electronic search strategies used controlled vocabulary and text words taking into account the characteristics of the population involved and the intervention and comparison used.<sup>(8)</sup> We did not specify the outcomes assessed in the truncated trials in the search strategies in order to

improve search sensitivity. We augmented our electronic literature searches with hand searches of bibliographies of eligible trials and personal communication with content experts.

### *Study selection*

We used online reference management systematic review software (DistillerSR, Evidence Partners, Ottawa, Canada; <http://systematic-review.net/>) to facilitate study selection. Pairs of independent reviewers screened the abstracts and then the full text version of potential references. Disagreements during abstract screening were included for full text screening and disagreements during full text screening were resolved by discussion or adjudication by a third reviewer. Agreement among reviewers on study selection using the kappa statistic averaged 0.71.

### *Data collection*

Reviewers first judged whether patients, interventions, comparators, and outcomes of each subsequent trial matched those of the corresponding truncated trial using the same criteria established in a previous study (STOPIT- 2, closeness rated on a 4-point Likert scale that ranged from 0-3).<sup>(3)</sup> Trials receiving the lowest similarity score for their population or intervention were excluded. Agreement among reviewers averaged 0.70 on the closeness criteria for patient population, interventions, comparators, and outcomes. Using standardized, pilot-tested data extraction forms and a detailed instruction manual, pairs of reviewers extracted data from eligible articles independently and in duplicate. We extracted subsequent trial characteristics (e.g. dates of enrollment and publication), whether a subsequent trial cited the corresponding truncated trial and

the rationale for launching or continuing the subsequent trial despite knowing the results of the truncated trial. We also extracted data on whether subsequent trials provided a sample size calculation, whether sample size calculation was informed by truncated trial results, whether a data monitoring committee (DMC) was available, whether truncated trial publication led to a change in the subsequent trial protocol (e.g., interim analysis, unblinding), and congruence between truncated trial and subsequent trial results with respect to the direction of effect and statistical significance.

Pairs of reviewers assessed risk of bias, independently and in duplicate, of truncated trials using items from the Cochrane risk-of-bias tool (9) focusing on allocation concealment, blinding, and loss to follow-up. For all phases of data abstraction and risk of bias adjudication, reviewers resolved disagreements by discussion or, if necessary, by third party adjudication.

#### *Outcomes of interest and statistical analysis*

The main outcome of interest was the incidence of subsequent trials (proportion of truncated trials that were followed by at least one subsequent trial). We conducted multiple Poisson and logistic regression analysis to explore the extent to which the following a priori established variables were associated with the decision to conduct a subsequent trial: sample size, number of events, funding source (non-profit/government vs. for-profit), allocation concealment (yes vs. no/unclear), patient blinding (yes vs. no/unclear), provider blinding (yes vs. no/unclear), presence of a DMC (yes vs. no/unclear), the explicit use of a stopping rule (yes vs. no/unclear), and time since publication of truncated trial (time from the publication year of truncated trial to 2015).

These variables were explored as potential confounders in regression. Poisson regression was used when the dependent outcome was the number of subsequent trials. Multiple logistic regression was used when the dependent outcome was “no subsequent trials published” vs “at least one subsequent trial published”. We conducted sensitivity analysis to evaluate truncated trials published later than 1990. We hypothesized using this arbitrary date as a cutoff that the reporting and conducting of randomized trials have improved over time; particularly after guidelines for reporting trials have been implemented by many journals. A two tailed p-value  $<0.05$  was regarded as being statistically significant. All statistical analyses were conducted using Stata version 14.0 (StataCorp LP, College Station, TX).

## RESULTS:

The literature search yielded 3,217 potential subsequent trials, of which 262 proved eligible (figure 1 of the supplement). Overall, reviewers judged subsequent trial questions to be similar to truncated trial questions. On a 4-point scale where 3 is most similar, similarity or closeness scores of the patient, intervention, comparison and outcome averaged 2.25, 2.39, 2.52 and 2.50; respectively. Subsequent trials were published on average 4 years after corresponding truncated trials. Table 1 of the supplement includes the bibliography of all subsequent trials and related truncated trials with year of publication, primary outcome and closeness scores.

### *Incidence and characteristics of subsequent trials:*

Overall, 102/207 (49%; 95% confidence interval 42%-56%) of truncated trials were followed by at least one subsequent trial addressing a similar question. The majority of truncated trials that were followed by subsequent trials had more than one related subsequent trial (56/102, 55%); median 2 (range 1-13). The characteristics of truncated trials are provided in table 1. The most common clinical areas addressed in these trials were cardiology (25%), cancer/malignancy (21%), human immunodeficiency virus infections (12%) and critical care (7%).

**Table 1 Description of 207 trials stopped early for benefit**

|  | Truncated trials without subsequent ones (N=105) | Truncated trials with subsequent ones (N=102) |
|--|--|---|
| Mean sample size                                     | 965 (range: 8-19,257)                            | 1,330 (range: 12-22,071)                      |
| Allocation concealment                               | 49.5%  | 52.9%   |
| Patient blinding                                     | 50.5%  | 52.9%   |
| Provider blinding                                    | 40.0%  | 42.2%   |
| Funding by nonprofit/government sources              | 28.6%  | 39.2%   |
| Presence of DMC                                      | 64.8%  | 68.6 %  |
| Explicit use of stopping rule                        | 68.6 %   | 77.5%   |
| Time since the publication of truncated trial (Year) | 15 (range: 8-27)                                 | 18 (range: 8-50)                              |

Although the majority of subsequent trials (179/262, 68%) cited the truncated trial and most subsequent trials 152 (58%) presented their results in the context of existing evidence (i.e., summarized research from other trials in their discussion section), only 99 (38%) provided a rationale for proceeding with an existing trial or launching a new trial in the face of the prior truncated trial. Reasons for continued investigation included indirectness (i.e., to test the intervention in a similar but different population or setting) 41/99 (41%), the need to collect data on other outcomes (21/99, 21%), or due to skepticism regarding the truncated trial results because of small sample size (12/99,

12%), increased risk of bias (7/99, 7%), or inconsistency of the results of the truncated trial with other evidence (11/99, 11%). The characteristics of subsequent trials are provided in table 2.

**Table 2 Description of 262 subsequent trials**

|   | No. | %     |
|---|-----|-------|
| <b>Subsequent trial cited the corresponding truncated trial*</b>            | 179 | 68.3% |
| Introduction  | 140 | 78.2% |
| Methods   | 30  | 16.8% |
| Results   | 9   | 5.0%  |
| Discussion  | 148 | 82.7% |
| <b>Rationale for launching or continuing randomization</b>                  | 99  | 37.8% |
| Subsequent and truncated trials have different questions*                   | 41  | 41.4% |
| -Different population   | 32  | 32.3% |
| -Different Intervention   | 11  | 11.1% |
| -Different way to measure the outcome                                       | 2   | 2.0%  |
| Need for more data on other outcomes  | 21  | 21.2% |
| Need for more data about harms  | 0   | 0.0%  |
| Insufficient sample size or number of events in the truncated trial         | 12  | 12.1% |
| Truncated trial results are inconsistent with other trials                  | 11  | 11.1% |
| Risk of bias of the truncated trial   | 7   | 7.1%  |
| <b>Impact of truncated trial publication on conduct of subsequent trial</b> |     |       |
| Sample size calculation based on truncated trial                            | 8   | 3.1%  |
| Stopped early   | 7   | 2.7%  |
| Interim analysis but not stopped  | 11  | 4.2%  |
| <b>Congruence of the results between truncated and subsequent trial</b>     |     |       |
| In the same direction and both are statistically significant                | 109 | 41.6% |
| In the same direction but subsequent trial not significant                  | 88  | 33.6% |
| In different directions and subsequent trial not significant                | 29  | 11.1% |
| In different directions and subsequent trial is significant                 | 2   | 0.8%  |
| Unable to determine because outcomes reported differently                   | 34  | 13.0% |

\* Categories are not mutually exclusive

In general, subsequent trials did not consider results of a truncated trial for sample size calculation or analysis plan. Only rarely did the publication of a truncated trial lead to the subsequent trial undergoing an interim analysis, breaking randomization codes or stopping early (4% of subsequent trials conducted interim analysis and 3% were



stopped early). Thirty-nine percent of subsequent trials reported using a DMC and 58% presented an explicit sample size calculation.

The results of subsequent trials were compared to those of truncated trials in terms of the outcome used to justify early stopping in truncated trials. The treatment effect favored the opposite intervention compared to the truncated trial in 12% of the trials (statistically significant in 1%, and non-significant in 11%).

An illustrative example of a truncated trial and corresponding subsequent trials is presented in Table 3.

**Table 3 An illustrative example of a truncated trial and subsequent trials\***

| Trial  | Sample size<br>(number of<br>events) | Effect on<br>mortality<br>(relative risk and<br>95% CI) | Trial rationale   |
|--|--------------------------------------|---|---|
| 2001: A randomized trial compared intensive insulin therapy to conventional treatment in critically ill surgical patients with raised serum glucose and was stopped early for significant reduction in mortality(10) | 1,548 (35 vs 63)                     | 0.58 (0.38-0.78)  | Trial was stopped early for benefit (index trial)   |
| 2006: A trial randomized critically ill patients in a medical ICU to intensive insulin therapy or conventional treatment(11)   | 1,200 (144 vs 162)                   | 0.93 (0.81-1.08)  | Investigators wanted to test the intervention in a different population (non-surgical patients) |
| 2008: A trial randomized critically ill patients   | 537 (61 vs 75)                       | 0.95 (0.71-1.27)  | Investigators wanted to test the intervention in a different population (severe sepsis).        |

|  |                    |                   |   |
|--|--------------------|-------------------|---|
| with sepsis to intensive insulin therapy or conventional treatment(12)   |                    |                   | Trial was stopped early at first planned safety interim analysis due to increased hypoglycemic events   |
| 2008: A trial randomized critically ill patients in medical and surgical ICUs to intensive insulin therapy or conventional treatment(13) | 523 (72 vs 83)     | 0.84 (0.64-1.09)  | Investigators considered the evidence-base insufficient to support intensive insulin therapy for critically ill patients  |
| 2009: A trial randomized critically ill patients in medical and surgical ICUs to intensive insulin therapy or conventional treatment(14) | 6,104 (829 vs 751) | 1.11 (1.01- 1.23) | Unconvinced with prior evidence, investigators wanted to resolve uncertainty about benefits and harms of intensive insulin therapy for critically ill patients. |

Abbreviations: RR, relative risk; CI, confidence interval; ICU, intensive care unit

\*Additional subsequent trials on this topic exist; however, we only describe here trials with more than 500 randomized patients.

### *Poisson regression and logistic regression*

Poisson regression and logistic regression demonstrated a statistically significant association between the incidence of subsequent trials and the time since the publication of the truncated trial suggesting that the longer the time since the truncated trial was published, the more subsequent trials were published. We did not find statistically significant associations between publication of subsequent trials and all other pre-specified explanatory (independent) variables (characteristics of truncated trials such as risk of bias, sample size, number of events, funding, presence of a DMC, and the explicit use of a stopping rule) (Table 4).

**Table 4: Association between incidence of subsequent trials and characteristics of truncated trials**

| <b>Model 1: Poisson regression of the number of subsequent trials on truncated trial characteristics</b>        |                     |             |         |
|---|---------------------|-------------|---------|
|   | Incident Rate Ratio | 95% CI      | p value |
| Sample size (per 1,000 patients)  | 1.02                | 0.96, 1.09  | 0.55    |
| Number of events (per 1,000 events)   | 1.25                | 0.57, 2.73  | 0.57    |
| Allocation concealment  | 0.92                | 0.54, 1.56  | 0.76    |
| Patient blinding  | 0.98                | 0.51, 1.87  | 0.95    |
| Provider blinding   | 0.74                | 0.39, 1.40  | 0.35    |
| Funding by nonprofit/government sources   | 0.77                | 0.50, 1.19  | 0.24    |
| Presence of DMC   | 1.11                | 0.57, 2.19  | 0.76    |
| Existing stop rule  | 1.81                | 0.99, 3.31  | 0.06    |
| Time since the publication of subsequent trial (Year)   | 1.07                | 1.04, 1.10  | <0.001  |
| <b>Model 2: Logistic regression, the publication of any subsequent trial on truncated trial characteristics</b> |                     |             |         |
| subsequent trial  | Odds Ratio          | 95% CI      | p value |
| Sample size (per 1,000 patients)  | 1.03                | 0.90, 1.18  | 0.68    |
| Number of events (per 1,000 events)   | 1.38                | 0.19, 10.12 | 0.75    |
| Allocation concealment  | 0.98                | 0.48, 1.99  | 0.95    |
| Patient blinding  | 0.97                | 0.33, 2.86  | 0.96    |
| Provider blinding   | 0.79                | 0.26, 2.36  | 0.67    |
| Funding by nonprofit/government sources   | 1.02                | 0.48, 2.17  | 0.95    |
| Presence of DMC   | 1.68                | 0.72, 3.88  | 0.23    |
| Existing stop rule  | 1.66                | 0.74, 3.71  | 0.22    |
| Time since the publication of subsequent trial  | 1.11                | 1.04, 1.18  | <0.01   |

### ***Sensitivity analysis***

When we excluded truncated trials published earlier than 1990, the conclusions remain unchanged. Ninety /190 (47%; 95% confidence interval 40%-55%) of truncated trials were

followed by at least one subsequent trial addressing a similar question (supplemental tables 2 and 3).

## DISCUSSION

Previous empirical studies have demonstrated that the practice of stopping trials early for benefit is increasing over time, that many of these trials do not report a rigorous approach for the decision to stop, that many exaggerate treatment effects by up to one third, and that when sample sizes and number of events are modest, the overestimates may be very large.<sup>(3)</sup> In the current study we found that 49% of truncated trials were followed by subsequent trials. In only a minority of these did investigators provide a rationale for randomizing patients despite the prior truncated trial, and in only a minority of these did the investigators suggest that findings required replication in an appreciably different population or setting.

The strengths of this study include a systematic and extensive literature search performed by an experienced librarian and supported by several methodologists. Tasks in this review that required judgment were performed by pairs of reviewers with resolution of conflicts by discussion or if necessary by third party adjudication. For subsequent trials we determined authors' rationale for proceeding, and tested possible associations between the decision to proceed and the characteristics of the initial truncated trials. Results were robust to sensitivity analysis when older trials were excluded.

One limitation of this study is the difficulty identifying stopped trials in electronic databases. The National Library of Medicine added the MeSH term "early termination of

clinical trial” in 2010 and EMBASE added the term in 2011; thus, trials published before these dates did not have specific indexing terms. In addition, the reliability of these terms and consistency in using them have not been formally studied. It is also plausible that the grey literature (eg, trial registries and conference proceedings) includes additional subsequent trials. Therefore, the current analysis may have underestimated the occurrence of subsequent trials published after truncated trials. Adjudication regarding similarity between truncated and subsequent trials required judgment, which may be flawed. Our reviewers were, however, all trained clinical epidemiologists. We made all judgments in duplicate, and chance-corrected agreement was good. Meta-epidemiologic research is retrospective, depends on the clarity of reporting in the published literature and can be subject to publication bias. Many of the truncated trials were published over 20 years ago with different reporting standards, making some judgments more difficult than others.

The finding of approximately half of the truncated trials being followed by subsequent trials demonstrates that authors of subsequent trials required more evidence than that provided by truncated trials, and that funders and ethics committees agreed with the necessity for further evidence. For the remaining half of truncated trials, a number of possibilities exist. The truncated trial may have provided compelling estimates to the extent that no further research was needed on a particular topic. Another possibility, which is most concerning, is that truncated trials have caused a “freezing effect” in that no further trials addressing the same question could be conducted even if a need was perceived. What would have happened had subsequent trials been conducted in these

instances? The results of subsequent trials that were in fact conducted – showed that many subsequent trials produced different results from the original truncated trials. In 41% of the subsequent trials, authors of subsequent trials cited indirectness of the intervention or population as a reason to launch their new trials or continue ongoing ones. In other words, they were interested in testing how the intervention might work in a population with somewhat different characteristics or test how the intervention would work if it was modified (e.g., given at a different dosage). To the extent that this rationale is warranted, it suggests that there may not have been a problem with the initial truncation decision.

Other rationales - the small sample size and number of events of truncated trials and the need to obtain data on additional outcomes of importance to patients – indicate serious problems with the initial truncation decision.<sup>(15)</sup> In the 62% of subsequent trials in which no rationale was provided, the absence of an explicit hypothesis of indirectness suggests the rationale was skepticism that the initial truncated trial results were in fact definitive.

The results of this study should impact the development of future trials and the decision-making process followed by DMCs. The ultimate goal of a DMC in this context should be to share benefits of an effective treatment as soon as there is sufficient confidence in the magnitude of its effects, on both benefits and harms.<sup>(16)</sup> Sufficient confidence would translate into an ethical mandate to offer the intervention to all patients, and to no longer randomize patients to the possibility of not receiving the intervention. Importantly,

the current analysis specifically addresses trials stopped early for benefit and not trials stopped early for harm (ie safety/adverse effects). Harm accounts for 16% of the reasons for stopping trials before their planned completion date,(17) is ethically considered an acceptable rationale for halting a trial, and does not lead to concerns about exaggerating the estimates of efficacy.

In the subsequent trials we have identified, investigators, funders and ethics committees decided that randomizing patients not only continued to be ethical but, in a setting of constrained resources was worthy of additional investigation. The subsequent trial investigators, funders, and ethics committees therefore disagreed with the DMCs and investigators of the truncated trials that an ethical mandate existed that all patients must now receive the putative beneficial intervention.

The current findings, in concert with prior ones (2, 3, 5, 16, 18-20) indicate the need for a re-evaluation of the criteria used for stopping early for benefit. Given what we know about the risk of overestimating treatment effects and the current perceptions of researchers demonstrated in this study, we suggest that DMCs and investigators entertaining stopping a trial early for benefit should ask themselves the following question: do experts and front-line clinicians agree that the question is definitively settled, all patients should now receive the treatment, and further investigation is ethically unacceptable? If these stringent criteria are not met and to ensure obtaining the maximal possible information about all relevant outcomes (which itself is an ethical mandate), the trial should continue.

## REFERENCES

1. Zannad F, Gattis Stough W, McMurray JJ, et al. When to stop a clinical trial early for benefit: lessons learned and future approaches. *Circ Heart Fail* 2012 Mar 1;5(2): 294-302.
2. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *Jama* 2005 Nov 2;294(17): 2203-9.
3. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *Jama* 2010 Mar 24;303(12): 1180-7.
4. Bassler D, Ferreira-Gonzalez I, Briel M, et al. Systematic reviewers neglect bias that results from trials stopped early for benefit. *J Clin Epidemiol* 2007 Sep;60(9): 869-73.
5. Hughes MD, Freedman LS, Pocock SJ. The impact of stopping rules on heterogeneity of results in overviews of clinical trials. *Biometrics* 1992 Mar;48(1): 41-53.
6. Poldermans D, Boersma E, Bax JJ, et al. The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group. *The New England journal of medicine* 1999 Dec 9;341(24): 1789-94.
7. Bouri S, Shun-Shin MJ, Cole GD, Mayet J, Francis DP. Meta-analysis of secure randomised controlled trials of beta-blockade to prevent perioperative death in non-cardiac surgery [published online July 31, 2013]. *Heart* 2014 Jul 31;100(6): 456-64.
8. Prutsky GJ, Domecq JP, Erwin PJ, et al. Initiation and continuation of randomized trials after the publication of a trial stopped early for benefit asking the same study question: STOPIT-3 study design. *Trials* 2013;14: 335.
9. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions In: Collaboration TC, editor.; 2011.
10. Van Den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in critically ill patients. *New England Journal of Medicine* 2001;345(19): 1359-67.
11. Van den Berghe G, Wilmer A, Hermans G, et al. Intensive insulin therapy in the medical ICU. *New England Journal of Medicine* 2006 Feb 2;354(5): 449-61.
12. Brunkhorst FM, Engel C, Bloos F, et al. Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *New England Journal of Medicine* 2008 Jan 10;358(2): 125-39.
13. Arabi YM, Dabbagh OC, Tamim HM, et al. Intensive versus conventional insulin therapy: a randomized controlled trial in medical and surgical critically ill patients. *Crit Care Med* 2008 Dec;36(12): 3190-7.
14. Nice-Sugar Study Investigators, Finfer S, Chittock DR, et al. Intensive versus conventional glucose control in critically ill patients. *New England Journal of Medicine* 2009 Mar 26;360(13): 1283-97.
15. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol* 2014 Jun;67(6): 622-8.
16. Pocock SJ. When to stop a clinical trial. *Bmj* 1992 Jul 25;305(6847): 235-40.
17. Reem A, Schandelmaier S, Kalu Olu K, et al. Premature trial discontinuation often not accurately reflected in registries: comparison of registry records with publications. *J Clin Epidemiol* 2016.
18. Pocock SJ. When (not) to stop a clinical trial for benefit. *Jama* 2005 Nov 2;294(17): 2228-30.
19. Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials* 1989 Dec;10(4 Suppl): 209S-21S.
20. Zhang JJ, Blumenthal GM, He K, Tang S, Cortazar P, Sridhara R. Overestimation of the effect size in group sequential trials. *Clin Cancer Res* 2012 Sep 15;18(18): 4872-6.



**Key Message:**

About half of the randomized controlled trials stopped early for benefit were followed by subsequent trials addressing a similar question

Trialists of subsequent trials may have been skeptic about the decision to stop prior trials; thus pursuing their own trials about the same question

There were no characteristics of truncated trials that were associated with, or could predict launching subsequent trials

A more rigorous threshold for stopping early for benefit is needed